LDPose: Towards Inclusive Human Pose Estimation for Limb-Deficient Individuals in the Wild

Supplementary Material

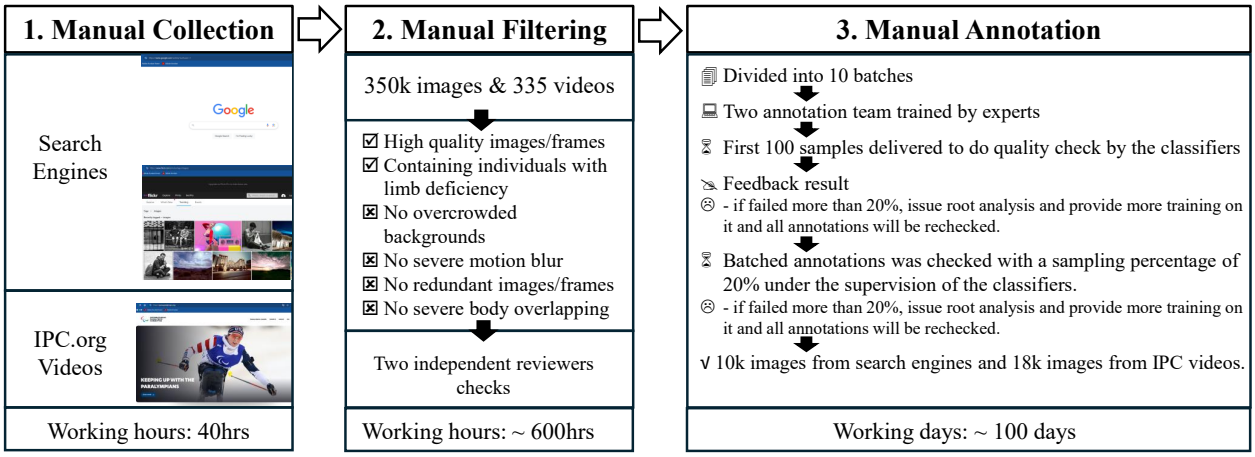

Figure 6. **Data collection and annotation process.**

9. Appendix

9.1. Boarder Impact

The introduction of LDPose, LDLoss, and LD Metrics represents a significant advancement toward inclusive and adaptive AI-driven pose estimation, with broad societal and technological implications. By addressing the long-standing exclusion of individuals with limb deficiencies from pose estimation research, this work promotes fairer, more representative models that can be integrated into assistive technologies, rehabilitation, and adaptive sports analytics. The dataset and methodologies introduced in this study provide new benchmarks for human pose estimation, encouraging future research into generalized models capable of handling diverse anatomical structures. Moreover, the proposed LDLoss and LD Metrics establish a new evaluation framework, ensuring that AI models are assessed not only on keypoint localization accuracy but also on their ability to differentiate intact and residual limbs correctly. These contributions have far-reaching implications, enabling advancements in prosthetic limb control, physical therapy, and accessibility-focused AI applications, ultimately fostering more equitable technological developments that empower individuals with limb deficiencies worldwide.

9.2. Data Collection, Filtering, and Annotation

The data collection, filtering and annotation process is demonstrated in Figure 6. After rigorous filtering and annotation, the final LDPose dataset consists of 10k images sourced from search engines and 18k images extracted from International Paralympic Committee (IPC) videos, ensuring a diverse and representative dataset for limb-deficient pose estimation. The entire data processing pipeline was highly labor intensive, requiring approximately 40 hours of data collection, 600 hours for filtering and quality assurance, and an estimated 100 days and 8 hours per annotator on 8 annotators for the annotation process. This systematic and quality-controlled approach guarantees high annotation accuracy, ultimately supporting the development of robust, fair, and generalizable pose estimation models for individuals with limb deficiencies.

Data Collection. The LDPose dataset was curated through a systematic multi-source data collection process to ensure diversity and representativeness in limb-deficient pose estimation. Images were acquired from search engines and the official International Paralympic Committee website, while video frames were extracted from 335 official IPC videos, capturing dynamic motion scenarios. The initial dataset were selected to encompass a wide range of limb deficiency types, environmental conditions, and activity settings.

Data Filtering. A rigorous manual filtering process was implemented to ensure the quality and usability of the dataset. Images were selected based on key criteria, including the presence of individuals with limb deficiencies to maintain dataset relevance, high-resolution images with clear body structures for precise keypoint annotation, and diverse environmental settings, covering indoor, outdoor, and competition environments to enhance generalization.

Table 3. Summary of Data in the LDPose Dataset.

Data Source	Total Images	Total Instances	Instances per Image	Resolution	Motion Blur	Environment Type
Online image search engines	10K	25K	1–2	200×143 to 8100×4752	None	Diverse (indoor, outdoor, stadium, urban)
Videos	18K	73K	1–17	480p to 1080p	Present	Highly crowded (stadiums, pools, race tracks)
Total (LDPose)	**28K**	**72.7K**	**3**	**High-variation dataset**	**Present**	**Diverse real-world conditions**

Conversely, images and frames were excluded if they exhibited overcrowded scenes, which could introduce annotation ambiguities, severe motion blur, compromising keypoint visibility, high redundancy due to similar or duplicate frames, and severe occlusions, where critical body parts were obscured. To ensure adherence to these filtering criteria, each image/frame underwent independent review by two annotators before proceeding to the annotation phase, guaranteeing consistency and accuracy in dataset selection.

Data Annotation. The annotation process was designed to be structured, systematic, and quality-controlled, ensuring accuracy and consistency throughout the dataset. To streamline the process, the dataset was divided into 10 batches, allowing for efficient management and verification. Each individual in the dataset was annotated with 25 keypoints, incorporating both standard body keypoints and newly defined residual limb endpoints, ensuring comprehensive pose representation. The initial quality review involved expert classifiers evaluating the first 100 samples to establish annotation consistency. If the failure rate exceeded 20 %, additional training sessions were conducted to refine the annotation process. Furthermore, each completed batch underwent batch verification, where a 20 % sample check was performed by internationally accredited para-athletics classifiers. If the failure rate for the batch exceeded 20 %, a full re-screening and correction process was initiated to maintain the highest standard of annotation quality.

9.3. Data Summary of Data in the LDPose Dataset

The web-sourced images consist of diverse resolutions and are captured in various real-world environments, featuring an average of approximately two individuals per image. The backgrounds vary widely, including hospitals, indoor training facilities, urban landscapes, beaches, and backyards, ensuring a rich contextual diversity. Additionally, the dataset encompasses a broad age range, with individuals ranging from approximately 5 to 90 years old, including both those with and without prosthetic limbs, further enhancing the dataset's inclusivity and representation of real-world scenarios.

In contrast, the video-extracted frames depict high-intensity sports performances, introducing realistic motion dynamics and challenges such as motion blur, occlusions, and crowded backgrounds. These frames contain 7 individuals per image on average, representing both athletes with limb deficiencies in the majority and able-bodied individuals in competitive environments. The dataset covers a range of adaptive sports, including para-athletics and para-swimming. The inclusion of varied weather conditions, complex background clutter, and severe occlusions ensures that LDPose captures realistic, in-the-wild pose estimation challenges, making it a robust benchmark for adaptive and inclusive AI models. More details about the data sources are listed in Table 3.

10. Dataset Demonstration

Figure 7. **Demonstration of LDPose Dataset**

Figure 8. **Demonstration of LDPose Dataset**

Figure 9. **Demonstration of LDPose Dataset**

11. Case Study

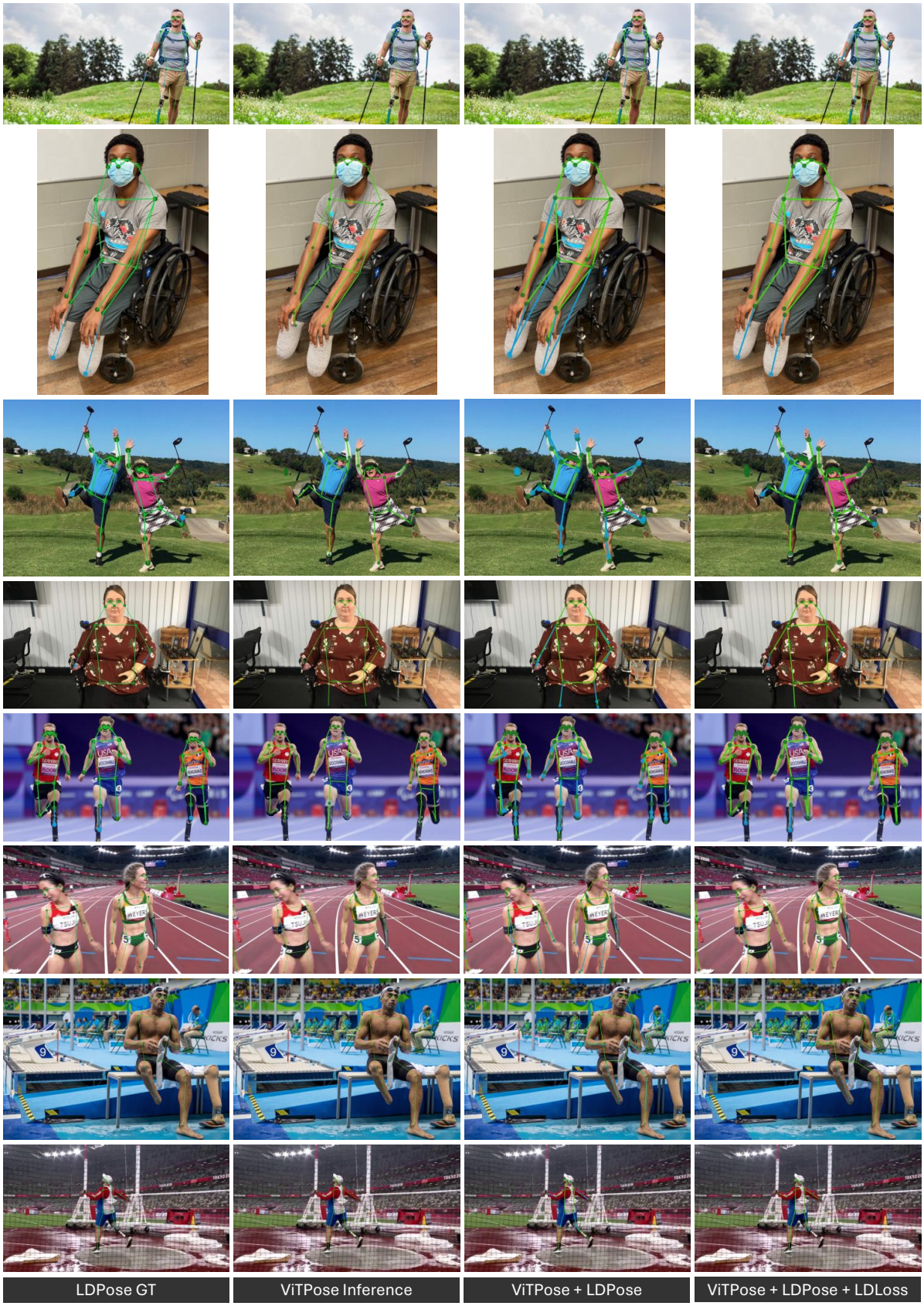

Figure 10. **More ViTPose with ViT-l backbone case study**